Exhibit M

**IN THE UNITED STATES DISTRICT COURT
FOR THE EASTERN DISTRICT OF VIRGINIA
ALEXANDRIA DIVISION**

United States of America, *et al.*,

    Plaintiffs,

    v

Google LLC,

    Defendant.

Case No. 1:23-cv-00108

HON. LEONIE H. M. BRINKEMA

EXPERT REPORT OF

GABRIEL WEINTRAUB, PH.D.

DECEMBER 22, 2023

"virtuous cycle."[247] Given the need for "sufficient" and the "right kind" of data for these algorithms, as outlined in this subsection, a company that falls out of the virtuous cycle will likely face difficulty in overcoming the hurdle when data scale, auction outcomes such as impressions won and revenue, and thickness are intertwined.

III.C. Scale and Experimentation

100. Technology companies routinely use experiments to assess the impact of changing product (or service) features or introducing new features. For these experiments, they often use A/B tests, also known as online controlled experiments ("OCE").[248] An A/B test is a randomized experiment in which a fraction of subjects are randomly selected to receive a treatment while the remaining subjects form the control group that represents the baseline (which is often the status quo). At Google, A/B tests are conducted using a tool called RASTA, which stands for "reliable AB stats tracking architecture."[249] Through RASTA and a system dubbed the "experiment framework", Google runs experiments on randomized samples of live web traffic[250] to evaluate

---

[247]   ▇▇▇▇▇▇-00003648, at -703 (12/2013) ("A virtuous cycle, network effects fuel more connections & data)", showing the "virtuous cycle" of more buyers leading to more data leading to more sellers leading to more data, etc.); *see* also, Deposition of ▇▇▇▇▇▇▇▇▇▇▇▇▇▇▇▇▇, 254:15–255:22 ("Q. Okay. This is gonna be a broad question. Other than what we just been discussing, can you explain the other ways in which scale matters in ad tech?…A.…In part, let me add, scale also affords data because, of course, in today's ecosystem, we are collecting vast sums of data and we are using that to inform our companies and our strategies and so with more scale comes more data and so that's a virtuous cycle."); *see also*, Deposition of ▇▇▇▇▇ ▇▇▇▇▇▇▇▇▇▇, 63:17–64:13 ("Q. Does ▇▇▇ vary its take rate by impression, meaning for a given publisher, ▇▇▇ might charge a different take rate for one impression than the other?…THE WITNESS: Yes. It varies – that varies by publisher in terms of how we apply our take rate, but a simple example is that we will often do something called fee squashing where we'll take a lower fee if we need to, to try and get an impression to run through our system. That would be an example.…Q. How, if at all, does greater scale affect exchanges' ability to effectively set different take rates for different impressions? A.…The greater the scale that you have, the more data you have. The more data you have, you can better decide what an appropriate fee would be on a given impression.").

[248]   Ron Kohavi, Diane Tang, and Ya Xu, *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing*, (Cambridge University Press, 2020), 5 ("Controlled experiments have a long and fascinating history, which we share online (Kohavi, Tang and Xu 2019). They are sometimes called A/B tests, A/B/n tests (to emphasize multiple variants), field experiments, randomized controlled experiments, split tests, bucket tests, and flights. In this book, we use the terms *controlled experiments* and *A/B tests* interchangeably, regardless of the number of variants.") (emphasis in original).

[249]   Deposition of Jason Hsueh (Google), November 15, 2023, 16:2–16:8 ("Q. What is RASTA? A. RASTA is the reliable AB stats tracking architecture. It's a tool to evaluate stats and impact of changes to our systems. Q. Does RASTA stand for reliable AB stats tracking architecture? A. Yes.").

[250]   Deposition of Jason Hsueh (Google), November 15, 2023, 23:3–24 ("Q. Does RASTA run experiments on live Web traffic? A. Yes. Q. Does RASTA pick samples from live Web traffic? A. Technically, it's a different

---

the impact of changes to Google's systems.[251] RASTA compares the effects of the experiment on the treated population to those of the control group, across various entities.[252] RASTA results are then linked to and summarized in launch documents which are reviewed by decision makers to determine if they should proceed with a given course of action, for example, whether to launch the treatment system-wide.[253]

101.   A/B tests are considered to be a scientifically valid approach to determine the causal impact of a treatment.[254] These randomized OCE's are the "gold standard for establishing causality,"[255] which is why web-facing companies including eBay, Facebook, Google, and Microsoft (among

---

system - experimentation system that does the sampling, but conceptually, yes. Q. What is the system that does the sampling?…A. It's an experiment framework. It's called the experiment framework. Q. What is the function of the experiment framework? A. It manages the configuration around changes and performs some of that random selection of different experiments. Q. Are RASTA experiments randomized? A. Yes.").

[251]   Deposition of Jason Hsueh (Google), November 15, 2023, 16:2–5 ("Q. What is RASTA? A. RASTA is the reliable AB stats tracking architecture. It's a tool to evaluate stats and impact of changes to our systems") and 18:21-19:1 ("Q.…So is RASTA used to evaluate the metrics for all changes within the ad serving system? A. Anything that's run as an experiment will typically go through RASTA.").

[252]   Deposition of Jason Hsueh (Google), November 15, 2023, 46:2–13 ("Q. In general, was Google able to use RASTA to look simultaneously at the results of a single experiment across the entities known at some times as DFP, AdX, and DB360 [sic]?…A. Yes, there are a number of views within RASTA that indicate - that impact on the various businesses…A. On the various products and businesses. It's part of our policies to understand the impact on various stakeholders.").

[253]   These launch docs are produced by a different tool, Ariane, which has since been replaced by a similar tool called Launch. *See* Deposition of Jason Hsueh (Google), November 15, 2023 16:9–19 ("Q. What is Ariane? A. Ariane is a tool for managing launches, the reviews, and the approval processes around those changes… Q. What is a launch? A. A launch is a change to products or systems that - a change to products or systems that we're making within Google.") and 19:2–14 ("Q. Does Ariane identify launches that include summaries of each launch? A. Yes. Q. Does each Ariane document also contain links to other documents and sources?…A. It may contain a variety of links within each ticket. Q. Does it link to experimental results contained in RASTA? A. For the launches that use RASTA, there would be links.") and 19:19–20:2 ("Q. Are there any databases or tools that came into use later in time after RASTA that serve a similar purpose? A. No. Q. Is the same true for Ariane? A. Ariane was replaced by the launch tool providing similar functionality. Q. What is a launch tool called? A. Launch.") and 58:2–18 ("Q. What is the difference between the two? A. A Design doc would typically be a document that states the problem and proposed solutions at the beginning of a project. A Launch doc is usually a recap of the project and the results.…Q. Do these kinds of documents typically get used in the normal course of Google's display advertising business? A. Yes.").

[254]   Ron Kohavi, Diane Tang, and Ya Xu, *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing* (Cambridge: Cambridge University Press, 2020), 9 ("Randomized controlled experiments are the gold standard for establishing causality.").

[255]   Ron Kohavi, Diane Tang, and Ya Xu, *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing* (Cambridge University Press, 2020), 9 ("Randomized controlled experiments are the gold standard for establishing causality.").

many others), use them to guide product development and accelerate innovation.[256] When designing experiments, there are best practices to ensure that the results are statistically valid; practices that Google aims to follow.[257] For example, Google takes measures to ensure the validity of their metrics, which include choosing random samples of web traffic and not informing participants that they are participating in an experiment.[258] Google also takes efforts to isolate each experiment from the others, and evaluates them independently so as to get more accurate results.[259] Included in the results of a RASTA experiment are "confidence intervals,"

---

[256] Ron Kohavi, Alex Deng, Brian Frasca, Toby Walker, Ya Xu, and Nils Pohlmann, "Online Controlled Experiments at Large Scale," *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2013): 1–9, at 1 ("Many web-facing companies use online controlled experiments to guide product development and prioritize ideas, including Amazon [1], eBay, Etsy [2], Facebook, Google [3], Groupon, Intuit [4], LinkedIn, Microsoft [5], Netflix [6], Shop Direct [7], StumbleUpon [8], Yahoo, and Zynga [9].").

[257] Deposition of Jason Hsueh (Google), November 15, 2023, 24:8–14 ("Q. Does Google try to design RASTA experiments to be as accurate as possible in a scientific method kind of way?…A. We try to make sure that we have statistical confidence within the metrics and that there is sound data behind it.").

[258] Deposition of Jason Hsueh (Google), November 15, 2023, 24:8–26:14 ("Q. Does Google try to design RASTA experiments to be as accurate as possible in a scientific method kind of way?…A. We try to make sure that we have statistical confidence within the metrics and that there is sound data behind it…Q. You mentioned before that RASTA experiments are randomized? A. Uh-huh. Q. Are there other design elements of RASTA experiments that are standardized that help ensure the experiment accuracy?…A. Not within RASTA. Q. In the experiment framework? A. In the experiment framework, that also remains largely within the random traffic selection. Q. In each RASTA experiment, is the traffic that receives the experimental treatment chosen randomly from the population that was eligible to participate in the experiment?…A. I didn't understand the difference between that and the question around traffic selection being random. Q. If your answer is the same, that's okay. A. Okay. I believe the answer is yes, that it's chosen randomly from the population. Q. Are RASTA experiments blinded?…A. I don't understand the term 'blinded.' Q. Are experiment participants aware that they are participating in an experiment? A. No.").

RASTA experiments typically use 1 percent to 5 percent sample sizes, and often include a ramp-up period. The ramp-up period is generally one to two weeks, but may be longer or shorter depending on the needs of the experiment. *See* Deposition of Jason Hsueh (Google), November 15, 2023, 26:14–19 ("Q. What sizes are the experiment group in RASTA typically? A. It may vary depending on the stage of experimentation. It frequently will ramp-up and start from less than 1 percent, and depending on the amount of data, it may go up towards 5 percent.").

A decision maker may choose to run an experiment for a longer period if they want to see more data, or if there was inconclusive evidence from the initial period, or may run it for a shorter period if they already have confidence with the results. *See* Deposition of Jason Hsueh (Google), November 15, 2023, 27:10–20 ("Q. Okay. Why would a decision-maker choose to run an experiment for a longer than standard period?…A. There might be a variety of reasons why a decision-maker would want to run an experiment longer. Q. Could you please name a few? A. Most likely if they wanted to see more data, if there was inconclusive evidence from the period run so far.").

[259] Deposition of Jason Hsueh (Google), November 15, 2023, 42:25–43:15 ("Q. Looking at the experiment, do the results measure only the impact of a single experiment or are there other changes in the auction that could be happening in other experiments simultaneously to the same group of impressions?...A. Yeah, I think it would be hard to tell from the data presented here. Q. Is it generally Google's practice to isolate each experiment from

---

which provide a statistical measure of Google's confidence in the results of the experiment as reported.[260] By utilizing these strategies, Google is able to test various ideas on a fraction of impressions or queries and learn with a high statistical level of confidence what the effect would be if applied to the total population.

102.   In this way, A/B tests have become an integral part of technological companies' innovation engines. These companies rely on experiments to gauge the effects of potential changes to their platforms before implementation.[261] Organizations use these experiments as a key tool for developing frameworks and testing ideas to understand the impact of the services they provide;[262] they have become a standard component of managerial decision making.[263] Companies like eBay, Facebook, Microsoft and Google run thousands to tens of thousands of experiments every year, testing everything from changes to the user interface, to customer

---

each other? A. Yes, we generally try to make sure that we have a mechanism for evaluating experiments independently. Q. Would you get more accurate results that way? A. Yes.").

[260]   Deposition of Jason Hsueh (Google), November 15, 2023, 28:3–10 ("Q. Is there something like a confidence interval that applies to experiments? A. Yes. Q. Can you describe it? A. The confidence interval is something that is serviced in RASTA, and it provides a statistical measure of our confidence within the results that are shown.").

RASTA defaults to a 95 percent confidence interval. *See* Deposition of Jason Hsueh (Google), November 15, 2023, 51:22–52:8 ("Q. Does 95 percent describe the confidence interval shown here? A. Yes. Q. Is that standard setting in RASTA? A. Yes, I believe so. Q. So if a RASTA experiment used standard settings, the 95 - excuse me, the confident interval would be 95 percent; is that right? A. I believe that's the typical case, but I haven't looked at specifically that, that field.").

[261]   Michael Luca, and Max H. Bazerman, *The Power of Experiments: Decision Making in a Data-Driven World* (Cambridge: MIT Press, 2021): vii-viii ("These days, companies like Google wouldn't dare make a major change in their platforms without first looking at experiments to understand how it would influence user behavior.").

[262]   Michael Luca and Max H. Bazerman, *The Power of Experiments: Decision Making in a Data-Driven World* (Cambridge: MIT Press, 2021), vii–viii ("From startups to international conglomerates to government agencies, organizations have a new tool to develop frameworks and test ideas, and to understand the impact of the products and services they are providing.").

[263]   Michael Luca and Max H. Bazerman, *The Power of Experiments: Decision Making in a Data-Driven World* (Cambridge: MIT Press, 2021), 62 ("But perhaps no sector has embraced the experimental method more than the tech sector, where it is now a standard component of managerial decision making.").

support systems, across websites, mobile apps and more.[264] For example, Google runs thousands of experiments across display ads.[265]

103.   A/B tests are often conducted with a relatively small yet representative sample. In display advertising, ad tech firms conduct A/B tests on a relatively small fraction of traffic before launching a new program or feature across their whole platform.[266] In this way, platforms reduce the potential risk of having a new feature produce an unexpected impact. For example, the new feature may degrade users' experiences, may reduce users' performance, or may involve an unexpected economic cost for the platform. The existence or magnitude of these negative effects may be unknown in advance; hence, a way of mitigating risk is to treat a relatively small fractions of units.

104.   Statistical power is an important criterion to consider when conducting statistically sound experiments.[267] Suppose an experiment is conducted to detect an X percent (or larger) improvement in an outcome of interest caused by a new feature. This "X percent" could be the smallest detectable effect that the company would use as a threshold to decide whether to launch the feature system-wide. Statistical power refers to the probability of detecting such an effect

---

[264]   Ron Kohavi, Diane Tang, and Ya Xu, *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing* (Cambridge University Press, 2020), 5 ("Online controlled experiments are used heavily at companies like Airbnb, Amazon, Booking.com, eBay, Facebook, Google, LinkedIn, Lyft, Microsoft, Netflix, Twitter, Uber, Yahoo!/Oath, and Yandex (Gupta et al. 2019). These companies run thousands to tens of thousands of experiments every year, sometimes involving millions of users and testing everything, including changes to the user interface (UI), relevance algorithms (search, ads, personalization, recommendations, and so on), latency/performance, content management systems, customer support systems, and more. Experiments are run on multiple channels: websites, desktop applications, mobile applications, and e-mail.").

[265]   Deposition of Jason Hsueh (Google), November 15, 2023, 23:25–24:7 ("Q. Circling back, how would you find out what the motivation was for performing an experiment?…A. I couldn't speculate for or state for all experiments. There's hundreds to a thousand, thousands of experiments that we run across display ads.").

[266]   Ron Kohavi, Diane Tang, and Ya Xu, *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing*, (Cambridge University Press, 2020): 171 ("In practice, it is common that an experiment goes through a *ramping* process to control unknown risks associated with new feature launches (aka. controlled exposure). For example, a new feature may start by exposing the Treatment to only a small percentage of users. If the metrics look reasonable and the system scales well, then we can expose more and more users to the Treatment. We ramp the traffic until the Treatment reaches desired exposure level.") *See, e.g.*, Deposition of Jason Hsueh (Google), November 15, 2023, 26:14–19 (Describing how Google display ad experiments typically use 1 percent to 5 percent sample sizes).

[267]   Ron Kohavi, Diane Tang, and Ya Xu, *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing* (Cambridge University Press, 2020), 30 ("Practically speaking, you want enough power in your experiment to be able to conclude with high probability whether your experiment has resulted in a change bigger than what you care about.").

when in fact one exists.[268] There are multiple factors that affect the statistical power of an experiment, such as the number of observations, where fewer observations lead to lower statistical power, and variance in the outcome of interest, where higher variance leads to lower statistical power.[269]

105.   A significant challenge in running A/B tests for technology platforms that run many experiments simultaneously is to secure enough users to achieve a desired level of statistical power. Technology companies are constantly launching, changing, and iterating on their products.[270] Each individual change may have a small effect, but the compounding effect of many small changes can be significant.[271] Therefore, detecting small effects reliably (or in a statistically significant way) is critical to be able to continuously experiment and improve.

106.   Statistical theory shows the size of the experimental sample should become larger as the level of the effect that an experimenter wants to detect becomes smaller.[272] For example, if a company

---

[268]   Ron Kohavi, Diane Tang, and Ya Xu, *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing* (Cambridge University Press, 2020), 30 ("*Statistical power* is the probability of detecting a meaningful difference between the variants when there really is one (statistically, reject the null when there is a difference).") (emphasis in original).

[269]   Ron Kohavi, Diane Tang, and Ya Xu, *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing* (Cambridge University Press, 2020), 30 ("Usually, we get more power when the sample size is larger."), and 189–190 ("Assuming Treatment and Control are of equal size, the total number of samples you need to achieve 80% power can be derived from the power formula above, and is approximately as showing in Equation 17.8 (van Belle 2008): $n \approx \frac{16\sigma^2}{\delta^2}$ where, σ^2 is the sample variance, and δ is the difference between Treatment and Control.").

[270]   Ron Kohavi, Diane Tang, and Ya Xu, *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing* (Cambridge University Press, 2020), 13 ("Features are built because teams believe they are useful, yet in many domains most ideas fail to improve key metrics. Only one third of the ideas tested at Microsoft improved the metric(s) they were designed to improve (Kohavi, Crook and Longbotham 2009). Success is even harder to find in well-optimized domains like Bing and Google, whereby some measures' success rate is about 10-20% (Manzi 2012).").

[271]   Ron Kohavi, Diane Tang, and Ya Xu, *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing* (Cambridge University Press, 2020), 14 ("[Google] ran hundreds of controlled experiments and multiple iterations; some across all markets, and some long term in specific markets to understand the impact on advertisers in more depth. This large backend change – and running controlled experiments – ultimately validated how planning multiple changes and layering them together improved the user's experience by providing higher quality ads, and improved their advertiser's experience moving towards lower average prices for the higher quality ads.").

[272]   Ron Kohavi, Alex Deng, Brian Frasca, Toby Walker, Ya Xu, and Nils Pohlmann, "Online Controlled Experiments at Large Scale," *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2013): 1–9, at 1 ("While running online controlled experiments require a sufficient number of users, teams working on products with thousands to tens of thousands of users (our general guidance is at least thousands of active users) are typically looking for larger effects, which are easier to detect than the

does not want to introduce a new feature unless an expected effect has a 1 percent lift on traffic, it should run a larger sample size experiment than if their target effect is 2 percent. As a result, large sample sizes are required for conducting reliable A/B tests for individual changes with relatively small effects.[273]

107. For this reason, it can be challenging for firms with limited scale to run informative experiments.[274] For technology platforms, the unit of an observation in an experiment typically corresponds to a user using the platform (e.g., a bidder), or to an event (e.g., a query, an auction or an impression). Hence, typically platforms with smaller scale need to run experiments over a longer time period before the number of users or events included in the experiment reaches the sample size required to detect the desired effect.[275] Running such long experiments can be impractical when a firm's objective is to iterate many small changes quickly. For example, most large tech companies run a single experiment for two weeks or less.[276] Consider another smaller company that has 20 percent of the traffic of the large tech company and uses the same treatment allocation fraction as the large company. Then, everything else equal, this smaller company

---

small effects that large sites worry about. For example, to increase the experiment sensitivity (detectable effect size) by a factor of 10, say from 5% to 0.5%, you need $10^2 = 100$ times more users.").

[273] Ron Kohavi, Diane Tang, and Ya Xu, *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing* (Cambridge University Press, 2020), 30 ("Usually, we get more power when the sample size is larger.").

[274] For example, LinkedIn reported that some experiments on its Learning Management Systems (LMS) have limited ability to detect revenue impact due to reduced sample size. LI_DOJ_G_2023_00000229, at -230 (10/2019) ("**LMS customer experiments have a limited ability to detect revenue impact**: In contrast to member-based experiments, LMS customer experiments have a low sample size and a high variance (in revenue) limiting our ability to detect revenue impact.") (emphasis in original); *see also*, "Find top products in Learning Management Systems (LMS) category," LinkedIn, accessed December 20, 2023, https://www.linkedin.com/products/categories/learning-management-systems.

[275] Ron Kohavi, Diane Tang, and Ya Xu, *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing* (Cambridge University Press, 2020), 33 ("**More users:** In the online experiments, because users trickle into experiments over time, the longer the experiment runs, the more users the experiment gets. This usually results in increased statistical power") (emphasis in original); *see also*, GOOG-AT-MDL-002207287, at -299 (06/2017) ("Starting an experiment…You may need to run longer or at higher traffic if confidence intervals are too wide").

[276] Somit Gupta, Ronny Kohavi, Diane Tang, Ya Xu, et al., "Top Challenges from the First Practical Online Controlled Experiments Summit," *ACM SIGKDD Explorations Newsletter* 21, no. 1 (2019): 20–35, at 21–22 ("There are many interesting and challenging open questions for OCE results analysis. While most experiments in the industry run for 2 weeks or less, we are really interested in detecting the long-term effect of a change… At Microsoft, while most experiments do not run for more than two weeks, it is recommended to run an experiment longer if novelty effects are suspected and use data from the last week to estimate the long-term treatment effect…At Twitter, a similar practice is followed. An experiment at Twitter may run for 4 weeks and data fromlast two weeks is analyzed.").

_____

Gabriel Weintraub, Ph.D.

December 22, 2023

---

that, and then we did not have the ability to invest on the people side. In the people side, technology businesses are fundamentally driven by people - whether that's writing code, whether that is designing a product, whether that is going into, as we say, a partner or a customer and selling - the core investment for growth that we typically make is around people. We did not have the resources after that to be able to make the investments in people.").

Expert Report of Gabriel Weintraub (December 22, 2023)--Errata

Page	Paragraph	Footnote	Original	Corrected	Reason
7	16	6	"ad**.**spend"	"ad spend"	Typo
9	19.b	.	"6.**94** percent"	"6.**24** percent"	Typo
9	19.b	.	"$**31** million"	"$**29** million"	Correction
11	22	10	"See, e.g., "Internet Advertising Revenue Report," PwC, iab, April 2023, https://www.iab.com/wpcontent/ uploads/2023/04/IAB_PwC_Internet_Advertising_Revenue_Report_2022.pdf"	"See, e.g., "Internet Advertising Revenue Report," PwC, iab, April 2023, https://www.iab.com/wpcontent/ uploads/2023/04/IAB_PwC_Internet_Advertising_Revenue_Report_2022.pdf, **at 15**"	Clarification
12	25	.	"The tools to run RTB auctions for the purchase and sale of online display ads **is** called…"	"The tools to run RTB auctions for the purchase and sale of online display ads **are** called…"	Typo
17	33	41	"(e.g., the average spend for each conversion should not exceed a pre-specified target).")."	"(e.g., the average spend for each conversion should not exceed a pre-specified target).") **(emphasis in original)**."	Clarification
19	34	47	"at 39**68**"	"at 39**52**"	Correction
23	43	71	"using machine algorithms"	"using machine **learning** algorithms"	Clarification
24	44	77	"no. 1 (**2014**)"	"no. 1 (**2015**)"	Typo
25	46	88	"EC '22: Proceedings of the **2023** ACM Conference on Economics and Computation **2023**"	"EC '22: Proceedings of the **23rd** ACM Conference on Economics and Computation"	Typo
25	46	88	"no. 1 (**2023**)"	"no. 1 (**2022**)"	Typo
25	46	90	"EC '22: Proceedings of the **2023** ACM Conference on Economics and Computation **2023**"	"EC '22: Proceedings of the **23rd** ACM Conference on Economics and Computation"	Typo
25	46	90	"no. 1 (**2023**)"	"no. 1 (**2022**)"	Typo
26	46	90	"(2023): 1–35, at 6"	"(2023): 1–35, at 6**–7**"	Clarification
26	46	91	"EC '22: Proceedings of the **2023** ACM Conference on Economics and Computation **2023**"	"EC '22: Proceedings of the **23rd** ACM Conference on Economics and Computation"	Typo
26	46	91	"no. 1 (**2023**)"	"no. 1 (**2022**)"	Typo
26	47	93	"See Lee Report, Section II.C.3. **and Section II.C.3.**, explaining…"	"See Lee Report, Section II.C.3., explaining…"	Typo
28	51	.	"Google Ad Manager ("GAM"), which includes a publisher ad **server, and** an ad exchange/SSP."	"Google Ad Manager ("GAM"), which includes a publisher ad **server and** an ad exchange/SSP."	Typo
32	59	132	"$\theta^1(x)$ "	"$\theta'(x)$ "	Typo
36	68	143	"("we use the *market thickness* to represent the average number of ads competing for user impressions on an online advertising platform.") **(emphasis in original)**;"	"("we use the *market thickness* to represent the average number of ads competing for user impressions on an online advertising platform.");"	Correction
37	69	.	"to **auctions with few participants** that"	"to **less competitive auctions** that"	Clarification
38	70	154	"**955**–1025"	"**965**–1025"	Typo
38	70	157	"**955**–1025"	"**965**–1025"	Typo
51	88	205	"things that we **think… helps** model highest other bid, and then we build a machine learning model."	"things that we **think helps** model highest other bid, and then we build a machine learning model."	Typo
54	92	218	"leaving money on the **table. We** also have access"	"leaving money on the **table. …We** also have access"	Clarification
56	93	225	"1849–1864, at **1850**"	"1849–1864, at **1851**"	Typo
60	98	243	"209–218, at **217**"	"209–218, at **209**"	Correction
60	98	244	"209–218, at **217**"	"209–218, at **216**"	Correction
63	101	258	"24:8–26:**14**"	"24:8–26:**6**"	Clarification
64	101	260	"the **confident** interval would be 95 percent"	"the **confidence** interval would be 95 percent"	Typo
64	102	261	"(Cambridge: MIT Press, **2021**)"	"(Cambridge: MIT Press, **2020**)"	Typo
64	102	262	"(Cambridge: MIT Press, **2021**)"	"(Cambridge: MIT Press, **2020**)"	Typo
64	102	263	"(Cambridge: MIT Press, **2021**)"	"(Cambridge: MIT Press, **2020**)"	Typo
66	104	269	"and 189**–190**"	"and 189"	Clarification

Expert Report of Gabriel Weintraub (December 22, 2023)--Errata

Page	Paragraph	Footnote	Original	Corrected	Reason
68	109	279	"at 3933"	"at 3933–**3934**"	Clarification
68	109	279	"(2017): 500–522."	"(2017): 500–522, **at 500, 510**."	Clarification
70	112	284	"**23:6**–24:6"	"**24:1**–24:6"	Clarification
70	112	.	"John Gentry of OpenX named personnel costs alongside infrastructure costs, and real estate costs as costs that need **to covered** in order to have the "ability to invest and grow the business.""	"John Gentry of OpenX named personnel costs alongside infrastructure costs and real estate costs as costs that need **to be covered** in order to have the "ability to invest and grow the business.""	Typo
71	113	286	"(09/09/2023)"	**No Date**	Correction
73	115	296	"247:17–248:**19**"	"247:17–248:**7**"	Clarification
84	130	313	"based on the analysis in **Professor Robin Lee's office**."	"based on the analysis in **the Lee Report**."	Clarification
95	140	.	"Indeed, since **2014**, Google consistently has a higher volume of impressions, higher win rate, and higher aggregate ad spending"	"Indeed, since **2015**, Google consistently has a higher volume of impressions, higher win rate **(since 2018)**, and higher aggregate ad spending"	Typo and Clarification
96	141	.	"since at least January **2015**"	"since at least January **2018**"	Typo
102	151	355	"GOOG-TEX-00034461, at -462"	"GOOG-TEX-00034461, at -462**–463**"	Clarification
102	152	356	"GOOG-DOJ-15564937, at -937**–938**"	"GOOG-DOJ-15564937, at -937"	Clarification
112	165		"James Avery of Kevel reported **to his colleagues** that"	"James Avery of Kevel reported that"	Correction
113	166	391	"at -095 (**02/01**/2021)"	"at -095 (**01/22**/2021)"	Correction
113	167	.	"Similarly, among ad networks, Google Ads' impressions and spend are more than 5 and 2 times larger than that of Criteo and FAN combined."	Similarly, among ad networks, Google Ads' impressions and spend are more than 5 and 2 times larger than that of Criteo and FAN combined **in many months**.	Clarification
114	169	.	"The **introduction** of Open Bidding by Google in 2018 allowed other exchanges to compete with AdX on a real-time basis, but with limitations."	"The **implementation** of Open Bidding by Google in 2018 allowed other exchanges to compete with AdX on a real-time basis, but with limitations."	Clarification
115	169	396	"189:5–190:**18**"	"189:5–190:**3**"	Clarification
120	182	.	"highest price"	"highest **historical average** price"	Clarification
120	182	.	"AdX could **still** use that competing header bid"	"AdX could use that competing header bid"	Clarification
124	188	434	"predicting bids, the parameter"	"predicting bids, **with** the parameter"	Typo
130	198	453	**Workpaper 9**, at tab Last Look.	**Conduct Effects Workpaper**, at tab Last Look.	Clarification
131	198	454	**Workpaper 9**, at tab Last Look.	**Conduct Effects Workpaper**, at tab Last Look.	Clarification
141	212	500	"GOOG-AT-MDL-B-001601271, at 273 ("<u>Sell-side dynamic revshare:</u> [l]aunched for +$**100**M revenue.") (emphasis in original)."	"GOOG-AT-MDL-B-001601271, at 273 ("<u>Sell-side dynamic revshare:</u> [l]aunched for +$**110**M revenue.") (emphasis in original)."	Typo
143	217	.	"For impressions, rival exchanges saw a decrease of 81.7 billion impressions, or 2.39 percent of annual impression volume **absence** the conduct."	"For impressions, rival exchanges saw a decrease of 81.7 billion impressions, or 2.39 percent of annual impression volume **absent** the conduct."	Typo
146	220	.	"Recall that in first price auctions, it is optimal for bidders to reduce their bids (Section **II.A.1**)."	"Recall that in first price auctions, it is optimal for bidders to reduce their bids (Section **II.C.1**)."	Typo
148	224	527	"improve **accuracy. Total** projected revenue loss"	"improve **accuracy... Total** projected revenue loss"	Clarification
152	229	552	"Deposition of ▮▮▮▮▮▮▮▮▮▮▮, 21:15–21:22 ("Q. How, if at all, did the decrease in spending by DV360 or ▮▮▮ affect ▮▮▮?...THE WITNESS: It was a devastating impact to the company, resulting in severe financial consequences. We had to ▮▮▮ ▮▮▮▮▮▮▮ and had a lot of negative effects as a result of that.")."	Switch with text in footnote 553.	Correction

Expert Report of Gabriel Weintraub (December 22, 2023)--Errata

Page	Paragraph	Footnote	Original	Corrected	Reason
152	229	553	"Deposition of ███████████████, 22:17–23 ("Our revenues went from - we were about ███ in 2017; 2018 we were about ██ because the first half of the year was strong; in 2019, we were down to about ████████; and in 2020, the combination of difficulties we had had plus COVID took us down to about ████████████ in net revenue in 2020.")."	Switch with text in footnote 552.	Correction
152	229	.	"The CEO of ██████████, testified that ██████ revenues fell from ███████████ in 2017 to ████████ 2019."	"The CEO of ██████████, testified that ██████s revenues fell from ███████ in 2017 to ████████ 2019."	Typo
154	232	564	"Poirot has actually been quite effective, resulting in DBM spending <u>7% more on AdX</u> and reducing spend on most other exchanges.")"	"Poirot has actually been quite effective, resulting in DBM spending <u>7% more on AdX</u> and reducing spend on most other exchanges.") **(emphasis in original)**"	Clarification
162	247	.	"introduction of **UPR** was attributable"	"introduction of **UFPA** was attributable"	Correction
165	252	.	"AdX's 6.48 percent gain in impressions translates to a loss of at minimum **50.9** billion impressions per year"	"AdX's 6.48 percent gain in impressions translates to a loss of at minimum **50.8** billion impressions per year"	Typo
191	290	690	"GOOG-AT-MDL-018528378, at -381 (03/02/2017) ("**Lost** of experiments!…Average ~120 per query")."	GOOG-AT-MDL-018528378, at -381 (03/02/2017) ("**Lots** of experiments!…Average ~120 per query").	Typo
C-1	3.a.ii.	3	"Letter from David R. Pearl to **Kelly Garcia**, September 8, 2023, 2."	"Letter from David Pearl to **Michael Freeman**, September 8, 2023, 2."	Correction
D-10	22	17	"Letter from Julie Elmer to John Hogan, August 19, 2022, 5–6|**8** ("	"Letter from Julie Elmer to John Hogan, August 19, 2022, 5–6 ("	Typo
D-12	28	25	"See GOOG-AT-MDL-008928566, at -583 (08/**03**/2017)"	"See GOOG-AT-MDL-008928566, at -583 (08/**23**/2017)"	Typo
E-2	2	.	"exchanges' bids **have the same variance** that is characterized **by the spread parameter**"	"exchanges' bids **have variance** that is characterized **by the same spread parameter**"	Clarification
E-21	42	21	"Letter from David Pearl to Michael J. Freeman, July 28, 2023 ("	"Letter from David Pearl to Michael J. Freeman, July 28, 2023**, 2** ("	Clarification
F-2	3	5	"John Rice, "11.2 Comparing Two Independent Samples", in Mathematical Statistics and Data Analysis, 3rd ed. (Duxbury: Thomson Brooks/Cole, 2007)."	"John Rice, "11.2 Comparing Two Independent Samples", in Mathematical Statistics and Data Analysis, 3rd ed. (Duxbury: Thomson Brooks/Cole, 2007)**, 421–444**."	Clarification
F-5	11	12	("Rule of Thumb[:] The basic formula is n=16/Δ^2 (2.3).")."	("Rule of Thumb[:] The basic formula is n=16/Δ^2 (2.3) **where Δ=(μ_0-μ_1)/σ=δ/σ (2.4)**.")."	Clarification
F-5	11	12	N/2=16 (p_p (1-p_p))/(p_1-p_0)	N/2=16 (p_p (1-p_p))/(p_1-p_0)**^2**	Correction
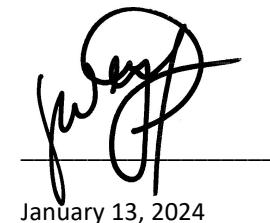
January 13, 2024

Expert Report of Gabriel Weintraub (December 22, 2023)--Supplemental Errata

Page	Paragraph	Footnote	Original	Corrected	Reason
142	216	509	This is close to the estimate we use from GOOG-DOJ-AT-02427435, at -435 of **6.58** percent.	This is close to the estimate we use from GOOG-DOJ-AT-02427435, at -435 of **8.6** percent.	Clarification

_____
February 23, 2024